

Machines and the face of ethics¹

Niklas Toivakainen
University of Helsinki

Introduction

Can machines become intelligent conscious agents? This is what one might call the classical philosophical question concerning Artificial Intelligence (AI) systems and robotics. But the space AI- and robotics technologies occupy in our society today, and the future visions attached to them, makes this classical question insufficient. For as these technologies become increasingly “autonomous” and as they start performing socially and interpersonally sensitive tasks, we become not only interested in their alleged or possible intelligence or consciousness, but in how they relate to ethics. An “autonomous” self-driven car or an “autonomous” military machine is not something we wish to — or so it seems — model simply on an intelligent or conscious agent, but on something that has some resemblance to an agent with a conscience. The same obviously applies for any machine that we wish to integrate as an “autonomous” actor in society, e.g. within health- and other services. Our society has also been eager to devise and construct technologies that will provide company, in one form or another, to people. Children as well as elderly people have more or less been the prime targets, although and especially within the sex industry, adults seek and are provided with artificial companions.

That is to say, the philosophy of AI and robotics has at least to some degree started witnessing what could be called an ethical turn, as the closer both the vision and the practice of AI and robotics come to actual human social life, the more the notion of and concern for “intelligence” or even “consciousness” becomes interdependent with a concern for ethics. One way of putting it is that this, in a sense, shows that our concepts of “intelligence” and “consciousness”, as they are used in the context of an actual human life, are always in relation to ethics.

Expanding the moral community?

In her recent book *Alone Together* Sherry Turkle describes how she became surprised and shocked when during an interview with *Scientific American* about “robots and our future”, she received the following accusation from the interviewer: “During that conversation, he accused me of harbouring sentiments that would put me squarely in the camp of those who have for so long stood in the way of marriage for homosexual couples. I was stunned, first because I harbour no such sentiments, but also because his accusation was prompted not by any objection I had made to the mating or marriage of people. The reporter was bothered because I had objected to the mating and marriage of people to robots” (Turkle 2012, 5).

“Exclusivist” was probably what the interviewer thought about Turkle — who claims that robots cannot be moral agents, and that any moral patiency ascribe to them is one of transference (ibid.). And for sure, the history of human culture (at least western culture) and its ethics has been modelled upon what could be called an exclusivist or centrist logic (cf. Gunkel 2012, Introna 2014) of the “we” and the “outsiders” or “them”, where the ethical project, as it were, has been to expand (and sometimes to diminish) the boundaries of the “we” in order to include — and thus to exclude — who or what is to be counted as an ethical agent and/or patient².

Notably the centrist logic not only captures something about how humans *have* acted, but reflects a quite generally shared moral self-understanding in today’s society. It also seems to be how Turkle’s interviewer understood the situation, calling for yet another expansion

¹ This paper was published in *Ethics and Information Technology*, Springer, 2015. DOI 10.1007/s10676-015-9372-y. The final publication is available at link.springer.com

² While moral agency focuses on whether an entity can act morally, moral patiency asks whether the entity is capable of, in some sense, suffering a loss essential to its self-integrity and being.

of the “we”, this time in order to include robots. And what he was calling for was of course not simply that humans should have the legal right to mate and marry robots, but that there is or should be a sense in which the robots “themselves” should be considered as at least moral patients, i.e. as beings who should be met with moral respect or consideration. Seen from this standpoint, the “machine question”, to borrow a phrase from David Gunkel (2012), looks much like the animal-rights or environmental-rights question.

As is simple enough to note the logic of exclusion is always based on some denominator or characteristic that separates “us” from “them”, while the task of inclusion is to either show how that feature is found also in the “other” (or a segment of them) or then to identify some new characteristic upon which the new “we” gets founded. “We are all human”, “we are all inhabitants of this planet”, “we are all sentient beings”. In the era of information technology, in the era when sophisticated machines become dominant and integrated parts of society, it is but ‘natural’ — or at least a fact — that the Universalist ethical project tries to keep up to date. In addition to placing moral patiency at the centre of ethics — as opposed to moral agency — Information Ethics (IE) (Floridi 2013), as the name suggests, holds that “fighting information entropy is the general moral law to be followed” (Floridi 2002, 300). In other words, IE, as a token of one of the latest versions of centric ethical universalism, suggests an approach where anything that is a coherent body of information ought to be taken as worthy of our moral consideration, something that obviously makes room for machines to enter into the domain of ethics.

But as critics of centrist ethics point out, the main problem with centrist ethics is not that it has a limited definition of the ethical domain — for could it be more Universal than IE? — but rather that the way in which it defines the ethical is determined by a subject or subjects (a group) that see themselves as privileged and standing at a place from which they can issue and ascribe ethical agency or patiency. As David Gunkel, who has sought for new forms of thinking about ethics in order to liberate what he calls the “machine question” (Gunkel 2012) from the constraints of centrist ethics, writes:

It [IE] is a problem insofar as this approach continues to deploy and support a strategy that is itself part and parcel of a totalizing, imperialist program. The problem, then, is not which centrism one develops and patronizes or which form of centrism is more or less inclusive of others; the problem is with the centrist approach itself. (ibid., 156-157).

So is the centrist, Universalist way of relating to ethics how one must think about ethics, or even, how one should think about it: does the centrist approach really capture the *meaning* of ethics? Although it is certainly fair to say that the centrist understanding of ethics has dominated at least the modern era, the history of philosophy up until this day has nevertheless witnessed thinkers that built their philosophical self-understanding on a very different notion of ethics, one in which ethics is, as it were, constitutionally entangled with philosophy or thought itself: ethics does not stem from reflecting on ethical principles, but finds itself always and already in the very act of thinking itself³. One of the leading 20th century proponents of this kind of philosophical self-understanding was the French philosopher Emanuel Levinas who developed his thinking in critical dialogue with the so called continental or phenomenological tradition, and especially with Heidegger’s version of it. Levinas’ as well as Heidegger’s philosophy has in recent times been taken up in connection with the philosophy of artificial intelligence (AI) and robotics and used to challenge the way in which ethics is thought about. One of the merits and hopes ascribed to this “new” way of looking at ethics has been that it promises a horizon of thought that overcomes or disentangles some of the classical problems within AI and robotics, problems that have suggested and threatened with definitive limits to artificial moral agency and patiency.

In what follows I will try to clarify the radical aspects of Levinasian ethics/philosophy and how it challenges the centrist approach. From this I will move on to give a

³ I am not here referring to any unified tradition in western philosophy. Nevertheless, as far as I can see, there have been many thinkers who have conceived of philosophy itself as always and already not only an ethical task, for to such a conception many would ascribe themselves, but more importantly to an understanding of thinking itself as ethically charged and even constituted. Some of these thinkers would include Socrates (see Wallgren 2006), St. Augustine and the early Christian thinkers, Nietzsche, Kierkegaard, Levinas and Wittgenstein.

critical account of David Gunkel's thoughts, who tries to develop an ethics of the machine on the basis of Levinasian philosophy. In connection with this critique I will also comment on the writings of Lucas Introna, who, while giving reference to Levinas, builds an idea of ethics based on Heidegger's analysis of the ontology of things. I will then proceed to articulate my critique of Introna and how the ethics he develops risk avoiding self-critical moral considerations central to our understanding of our relationship to artefacts and hence technology.

An ethics of the face

What does it mean to say that Universalist normative ethical theories such as IE still keep to the centrist approach? And what is the problem with it? In what sense do centrist ethics fail to give us an understanding of the *meaning* of ethics? One of the central questions that Levinasian philosophy brings to the forefront is how we are to understand, as one might put it, what it is that makes our ethical norms and principles *ethical*. Take for instance the normative claim of IE as articulated by Floridi: "fighting information entropy is the general moral law to be followed". As with any norm, the norm *itself* does not clarify what makes it ethical, but always points to something that is always already there, something in relation to which the norm is established, something that makes it *ethical*. There is, as Gunkel rightly notes, "a kind of morality before morality" (Gunkel 2012, 175), i.e. any ethical theory or norm already presupposes that which makes ethics *ethics*.

Obviously centrist ethics have tried to account for this, in various ways. Transcendentalism (Kant, 2002) or historicism (Westerlund, 2014) are examples that present themselves as viable and — to some — attractive options. But more in tune with our time are the aspirations to either explain morals by way of social construction (Hitlin & Vaisey, 2010) or by naturalising ethics by for instance arguing for a biological and evolutionary foundation for ethics (Joyce, 2001): ethics has fitness value for the human organism and has proven to be of advantage in the evolutionary struggle for survival. Crucially, what makes all such attempts to articulate ethical norms and their foundations centrist, is the very notion that ethics can be *determined* by rational abstractions and theories and is in this sense something *we* can *decide* on and *control*. As Lucas Introna comments, in centrist ethics "we have already chosen, or presumed, the framework of values that will count in determining moral significance — that is, who is in *our* circle and who is outside of it, and for what reasons" (Introna 2014, 45).

Centrist ethics has thus been criticised for its pursuit of *sameness*: first we have for instance the free men of Athens, then we start including others *into* the same: they become *like* "us", and it is *only* by becoming like us (or being left out) that they acquire their moral status. Even in the Universalist form of ethical centrism the main logical or rather *relational* structure remains the same. That is to say, it rests on a *picture* of a kind of privileged subject (or group) that begins with confirming its self-identity, determining the perspective from which ethical recognition and respect then flows out and is ascribed to an *other* (or group, limited or Universal) and thus setting the *ethical relationship*. In this picture then, the '*I*' and the *other* are there, as it were, before and independently of each other, self-sufficiently independent of their relationship to each other — like the Cartesian subject or the Hobbesian ego-centric individual.

As a radical alternative, and as a way of overcoming the centrist approach as well as addressing the issue of "a kind of morality before morality" — i.e. addressing how it is that others matter (ethically) to us at all — many thinkers, David Gunkel among them, have turned to the philosophy of Emmanuel Levinas⁴. Instead of the centrist picture of the self-constitutive and -determined subject looking out into the world and relating others to itself, Levinasian philosophy proposes that the *relationship* between the self and the other is in fact the constitutive foundation of the very structure of the subject and of ethics (Levinas, 1969 & 1985). As Gunkel notes:

⁴ Although I allude to Gunkel's reading of Levinas, the account given in this paper will depict Levinas in a slightly different light than Gunkel's interpretation. Some things I say will also be extensions or modifications of Levinas' thoughts.

The self or the ego, as Levinas describes it, does not constitute some pre-existing self-assured condition that is situated before and is the cause of the subsequent relationship with an other. It does not (yet) take the form of an active agent who is able to decide to extend him- or herself to others in a deliberate act of compassion. Rather it becomes what it is as a by-product of an uncontrolled and incomprehensible exposure to the face of the Other that takes place prior to any formulation of the self in terms of agency (Gunkel 2012, 177).

Ethics in Levinasian terms is uncontrollable and inescapable — not something the ‘self’ can determine — because, as Levinas puts it, the “face is meaning all by itself” (Levinas 1985, 86) — although it is always in relationship to another: It is not we that give the other’s face its meaning or construct it, but rather, it is through our *encounter* with the (meaning of the) face of the other that we become (gain the meaning of) a ‘self’, a self with thoughts, emotions, consciousness etc., a self that *then* can impose and project all kinds of ‘meanings’ on the other. Hence Levinas’ claim that ethics is first philosophy (Levinas 1969).

To claim, as Levinas would, that ethics, that the claim of the other, is uncontrollable and inescapable, is of course not to deny that our lives with each other is filled with attempts to control and determine ethics; to set norms and principles. Rather, the point here is that all our responses to others are always already in relation to (the meaning of) the ‘face’ of the other. This suggests that ethics is not about constructing theories or expanding ethical norms (although such can be done in relation to the pervasive ethical claim of the other), but rather always stands in relation to a dynamics of opening up to the face of the other — letting the face touch us without restrictions — contra escaping it and suppressing/repressing how the other touches us and how we respond to the other.

One might try to bring out the Levinasian idea with the help of an example. Imagine seeing a person on the street, say, crying desperately. Suddenly everything one does, everything one thinks about and feels will, in some sense, be related to this person. In other words, one cannot but respond to the cry. This is obviously not to say that everyone will respond by acknowledging the other and caring for him/her, for as we all know, we avoid, ignore, overlook etc. each other quite systematically, not only when going about our everyday lives with each other, but also in times of suffering and joy. Nevertheless, our very avoidance, indifference, lack of empathy etc. all stand in relation to that which is ignored, avoided, made indifferent etc. — all are, in their own right, affective responses to the cry of the other. So instead of approaching the desperately crying person — something that the desperation of the other claims of us, or as one might say, which is the meaning of the crying face — one might turn away or simply walk by indifferently. One might for instance reason to oneself that someone else is bound to come to help, that someone else will be in a much better position to care for the crying person. Or, one might reason that one is in such a hurry that one does not have the time to attend to the other (and that someone is bound to do so sooner or later), or that the person desperately crying is in some sense personally responsible for his/her misery — perhaps he/she appears to be homeless and alcoholic — and does not deserve to be cared for. These types of justifications need of course not be consciously articulated by anyone, but can have the form of unarticulated and unconscious motivations, which show themselves in the very way in which one responds/behaves towards the other — part of our everyday conventions. Nevertheless, the important point here is that all our responses will in some way or other be related to the person desperately crying. This we cannot escape.

It is important to note at this point that the characterisation of ethics in Levinasian terms is neither normative nor empirical. It is not normative in the sense that someone could determine how one ought to be or act. In the example of the desperately crying person, there is no collective (for instance our society) or subject (the crying person him-/herself or, e.g., one’s parents) as such that determines how one ought to respond to a crying person. It is the crying face, the meaning it has in itself and our unavoidable responsiveness to it that calls upon us. Neither is Levinasian ethics based on or relative to empirical data. Empirical data shows us the ways in which we *do in fact* treat and respond to each other, not the *meaning* of those very ways. So for example, as some research results seem to suggest, when people are faced with tight time schedules they are prone (not) to aid persons in need of care, not so much based on their values,

but rather based on “how much time pressure they felt” (Berreby 2005, 5)⁵. Nevertheless, empirical data cannot of course tell us the *meaning* — the ethical significance — of what is going on in the experiments. For instance, it cannot tell us that ‘time pressure’ is in fact not an, ethically speaking, *neutral* ‘sensation’ one has or a *neutral* fact about one’s condition, but is essentially connected to all the different relations that give sense to the ‘pressure’. In the experiment in question, the experimental subjects were students of theology on their way to give lectures and were given different timeframes to reach the lecture hall. Why did they feel time *pressure*? Well, one might suggest, at least partly because they had a feeling of responsibility they felt they wanted to fulfil. Or perhaps partly because they felt it to be an honour and a good merit on their behalf to be asked to give a lecture and thought it important to be obedient to schedules. Be it how it may, in addition to the fact that time pressure is always in relation to one’s relationship to others and one’s own self-conception and -identity, the important point to be made here is that the ‘time pressure’ felt by the theology students inevitably also became related to the person they saw fainting on their way to the lecture hall. They need not, as said already, admit this, or even notice it consciously, but the point is that insofar as they actually saw a person faint, whatever ‘time pressure’ they might have felt immediately became related to this person in need of help: their felt ‘time pressure’ became an *excuse* — consciously or not — not to respond with care to the fainting person.

All these different meaning possibilities are of course not something an empirical experiment can come to establish. Rather, as Levinasian ethics wants to point out, empirical research and experiments already find themselves within a space of ethics; there is no ‘neutral’ stance from which such research and experiments can be done. For instance, the sense of the above mentioned experiment is possible only to the extent the setup already as such acknowledges that we live in an ethical universe; that others are beings with faces — in Levinasian terms; that there is something like ‘stopping to help someone’ — that there is an ethical sense to this; that it is already acknowledged what it would mean to help another being, which in turn gives the sense to the idea that students were under time pressure and used it as a reason for not helping etc. What the experiments can of course show is how few actually do stop to help, how unwilling we are to take responsibility for each other and how self-centred we are.

As I have tried to point out, Levinasian ethics does not dwell within a framework of ethical principles and theories, but is rather situated within a dynamics of openness towards and a turning away from the other. Ethics is not something we construct but rather something we learn to take responsibility for and articulate in its own terms. In other words, ethics is not so much a question of intellectual problems but rather a struggle with difficulties of facing each other and combating our own self-centeredness. This brings to mind something Wittgenstein wrote:

Tolstoy: a thing’s significance (importance) lies in its being something everyone can understand. — That is both true and false. What makes a subject hard to understand — if it’s something significant and important — is not that before you can understand it you need to be specially trained in abstruse matters, but the contrast between understanding the subject and what most people *want* to see. Because of this the very things which are most obvious may become the hardest of all to understand. What has to be overcome is a difficulty having to do with the will, rather than with the intellect” (Wittgenstein 1980, 17e)

Wittgenstein’s suggestion that the difficulty lies in “what most people *want* to see” is of course another way of saying that the difficulty of understanding is imbedded *in* what most people do *not* want to see. What is it we do not want to see? This question is obviously outside the scope of this paper, but as far as can be said here, what the Levinasian perspective can help us see is that what we seem to fear is precisely the *ethical*. That is to say, we fear to let ourselves respond openly to the ethical relationship, to the face of the other, which on the one hand lies at the very foundation of our subjectivity, but at the same time threatens or undermines our (the subject’s) self-centred self-determining aspirations. In Levinasian terms, the ethical relationship to the other is first and foremost characterised as responsibility, an infinite responsibility for the other with no guarantees of reciprocity or mutual recognition: ethics is unconditional (Levinas 1985, 95-101). In this sense

⁵ See also Darley & Batson (1973)

the ethical claim disregards any egoistic or self-centred demands and aspirations. Rather, insofar as the subject's own desires and interests as well as fears and things like laziness, say, come in conflict with the responsibility for the other — the claim of the face — and insofar as the subject is driven by self-interest, ethics *threatens* the subject's will to power and self-determination. Another aspect of this difficulty and resistance of the will is that in the ethical relationship one loses the power of control. In opening up to the other, in letting the face of the other touch oneself without resistance, one cannot determine or control the flow of things: it is not the 'I' who determines what the relationship is and what/how the self must be. Rather, what will be is always as open as 'I' and the other allow it to be — normative (centrist) ethics is a way of controlling and managing exactly this difficulty (cf. Toivakainen 2014 & 2015)⁶. It is in this light that I read Wittgenstein's remark "Nothing is so difficult as not deceiving oneself" (Wittgenstein 1980, 34e), for that which is the "most obvious", the closest of all, is that which threatens the "I" the most.

Approaching the question of the face of machines

As already mentioned, David Gunkel is one who has turned to Levinasian ethics in order to challenge and overcome classical problems within AI and robotics, problems that seem to set definitive limits to artificial moral agency and patiency. I shall now turn to give a critical account of Gunkel's attempt to apply Levinasian ethics to the "machine question", after which I will introduce the writings of Lucas Introna into the discussion and continue articulating my critique.

In characterising the nature of ethics Gunkel makes the following remark: "Before something is decided to be either a moral agent or a moral patient, *we make a decision* whether to make this decision or not" (Gunkel 2012, 175, emphasis added) and "*We*, individually and in collaboration with each other (and not just those others who we assume are substantially like ourselves), *decide* who is and who is not part of the moral community" (ibid., 214, emphasis added). Given that Gunkel bases his thinking on Levinasian ethics, his remark seems to me to be more misleading than true/clarifying. It is misleading to the extent it is meant to suggest that we have a power to "decide" what our ethics is about, what it means. For as far as Levinas is concerned, we can never as such decide on who is and who is not a moral agent or moral patient — who has a face: we cannot choose our morality. Ethics is not a choice we make, but rather a relationship or space of meaning we already find ourselves in. On the other hand, decisions are made in *relation* to this moral necessity that claims us. That is to say, we *can* decide to embrace or repel the other *only insofar* as the other already has a face, insofar as we already stand in an ethical relationship to the other. So it is not up to me to decide whether or not, say, a slave or a member of some other culture is a moral agent/patient. In this sense I can neither construct (theoretically, normatively) a face nor deconstruct it: the face uncontrollably calls upon me. What I can do is to avoid or turn away from any particular face and in setting up norms one can decide who one wants to exclude/include. In other words, if there is a decision to be made in ethics, as articulated by Levinas, it is a decision to acknowledge and open up to the other, or then suppress or repress the bearing the other has upon us.

Hence it is not clear in what sense Gunkel is referring to a decision in the quote above. My suspicion is that there is some ambivalence and confusion involved and that it connects directly to Gunkel's main project, namely to the question of the 'face' of the machine.

Despite his respectful relation to Levinas' ethics as first philosophy, Gunkel is troubled by what he finds to be an exclusively humanistic face of the other in Levinas. Following Jacques Derrida's (2008) critique of Levinas, Gunkel reminds us that the animal was left without a face in Levinas and that on those "rare occasions when Levinas does address himself directly to animals — when he is, it seems, not silent about the animal — it is in order to silence or dismiss them from any further considerations" (Gunkel 2012, 180). As Gunkel also notes, animals and even eco-systems have nowadays gained an ethical face with the help of a radicalisation of

⁶ Here it is clear that I part with Levinas, who is prone to ascribe the repression of the openness to the other and the introduction of norms and justice not so much to the dynamics I tried to shortly introduce, but to a dynamics that unfolds because there are many others, not simply the one I encounter *now* (Levinas 1985). I should also like to add that my own account is indebted to the writings of my colleagues Joel Backström (2007) and Hannes Nykänen (2002 & 2014).

Levinasian ethics (cf. Calarco 2008, Benso 2000). Nevertheless, as Gunkel sadly informs us: “Although other kinds of previously excluded others have been slowly and not without considerable struggle granted membership in the community of moral subjects — women, people of color, some animals, and even the environment — the machine remains on the periphery. It exceeds and escapes even the best efforts at achieving *greater inclusivity*” (ibid., 206, emphasis added). That the machine does not have a *face* (ibid., 196), that the machine has been left out, is to Gunkel one of the big theoretical shortcomings in modern ethics. Hence cybernetics is taken up as one of the most promising developments that gives “a radical, posthuman *theory* that deposes anthropocentrism and opens up thoughtful consideration to previously *excluded others*” (ibid., 201, emphasis added).

I emphasised the words “theory” and “excluded” here, because I find both of them problematic and moving away from the understanding of ethics that Levinas is, I think, pointing towards. The most obvious problem with Gunkel’s way of characterising his project of the “machine question”, as he calls it, is that although he eagerly tries to overcome the centrist paradigm, his understanding of the “decision” involved in ethics takes him back to an idea that it is up to our decision to *include* the machine. And this is directly related to his conception that in order to reach a higher level of inclusion, we are in need of a better *theory*. Although Levinas himself is somewhat ambivalent on the issue, it seems to me that his articulation of the face of the other as the foundation of ethics is not a *theory* about ethics, but more likened to a *description* of the meaning/significance of the encounter with the face. The “face is”, as Levinas puts it, “meaning all by itself” (Levinas 1985, 86). As already noted above, I understand this to mean that who or what has a face is not up to us to decide, there is no new knowledge (at least in any theoretical sense) that might give an other a face. The ethical task is rather to let the face touch oneself, to acknowledge how it already touches one, and not to come up with some new theories that make it *intellectually* justifiable for us to include new entities into our ethical universe.

One must keep in mind that questions concerning the moral status of women, slaves, animals (to extend the Levinasian perspective) etc. have been questions relative to specific cultural contexts, although the issues have been widely shared by more or less all (high-)cultures. For instance, contemporary animal rights questions have emerged as burning issues relative to the instrumentalisation/mechanisation of animals. Other cultures, who have not shared the mechanistic picture of life and have not had grim meat and milk production industries, have not had to ponder whether for instance animals could feel pain or not. For them, the question of the sentience of animals and their dignity was and is not an issue, although other (and very similar) concerns might be. Similarly the question of the moral status of slaves is a question that has a place only in a society that exercises and tries to legitimise slavery.

What we need to note here is that in for instance our modern culture, which tends to completely instrumentalise animals — especially cattle and other animals used solely for the purpose of food or clothing — people have simultaneously always kept pets and, importantly, had affective and loving relationships with them⁷. In other words, at the same time as animals are deprived of moral dignity (especially in industry) and excluded from moral consideration to a very high degree, people are in ethical relationships with them: at the same time as they have an ethical claim on us, we suppress or repress it in the name of, say, our meat consumption⁸. The same conflict can be seen in the case of slaves. While it seems to have been a common argument of for instance the North American slave-owners that a person of African origin could not feel pain in the way white people could, that they were sub-human and unable to think higher or more complex thoughts and were unable to (learn to) read, the slave-masters displayed behaviours that showed clear signs of sexual attraction and (perverted) affection, enjoyment of punishing and torturing (why would you do this *if* the slave did not have a similar inner life as you?), jealousy (for instance

⁷ One should also add that especially in small scale farming, farmers usually have had and still have more direct and even affectionate relationships to their animals — this is one reason why animals are for the most part not slaughtered by the farmer him-/herself, and in any case, the act of slaughter always has a bearing on one’s conscience.

⁸ My point here is not to advocate vegetarianism, although I am in sympathy with vegetarians. My point here is simply to say that we are in an ethical relationship to animals, and the nature of that relation (which I do not qualify here) will have its bearing on meat-eating.

a wife's jealousy for her husband's desire for a female slave) and fear — why was it strictly forbidden to try and teach slaves to read if they *per se* were unable to learn it! All of these characteristics have been well documented⁹ and to my mind show that the slaves always had a face — like any white European —, that this face always was at the core of the relationship between the slave-owner and the slave and formed the perverted and destructive behaviours and the dynamics of the social structures. The problem was not that the slave did not have a face; the problem was that the slave-society suppressed and repressed it.

The main point of all of this is that the inclusion of animals, slaves, etc. has not come about because the formerly excluded have now been given a face — there has been no choice to grant them a face nor any theory that has constructed a face for them (as if they lacked a face before). Rather, it is more accurate to say that their faces have now been acknowledged and permitted to be (socially accepted/acknowledged as) faces. Of course it is also true that much of this acknowledgement has been translated into all kinds of new normative systems and ethical principles that are taken to be justified by and derived from certain theoretical abstractions, and not from the very relationship itself. Nonetheless, this does not change the basic structure of Levinasian ethics, nor does Levinasian ethics try to deny how much of our ethical self-understanding is dominated by our tendency to understand ethics in centrist terms, i.e. by our fear to take personal responsibility for ethics.

Now I can perhaps more clearly articulate what it is in Gunkel that makes him fall back on a centrist ethics, despite his aspiration to develop a Levinasian ethics. The central issue that Gunkel is not able to account for or clarify, is the issue of why it is that we should at all be concerned with our moral relationship — or lack of it — to machines? And this question will not be answered by saying that machines are still excluded from the moral community. For despite the seeming self-evident nature of the question of machine ethics, such a question is never neutral. That is to say, it is not a question that presents itself to us unconnected to our very aspiration to ask such a question. Why do we ask it? Why are we concerned with how to relate to machines? Why does Gunkel feel it to be important to ask about the face of machines? Have we mistreated the machine, the way slave-owners mistreated their slaves, in ways that indicate that it in fact has a face that has been repressed?

Centrist ethics asks its questions *as if* neutrally, as if observing the world and ethical relationships from the outside, as if in a position to discriminate “objectively” who and what should be included/excluded. Gunkel's question is like the centrist one. It is a question that seems to appear out of a neutral perspective — self-evidently — from which one observes that all kinds of different beings have now been included into a moral family, except that the poor machine is still left outside — but why does he/we care, why should we? So what, one may ask, gives Gunkel's question — can the machine have a face — and his aspiration — the machine *should* have some kind of a face — its sense? Levinasian ethics suggests that ethics is always bound to the *relationship* to the other and flows from it. That is to say, we are always already involved (personally) in our ethical thinking, we do not stand in a neutral position observing an ethical landscape. So the question one should ask is: what is our relationship to machines really like? Do they have an ethical bearing on us — perhaps in a hidden and repressed way? And what is the quality and different dimensions of this relationship? If the machine is to have a face, then it already has one, or if it is a question of a future machine then similarly it must bear its own ethical meaning and cannot be given one.

In search of the machine-face

If one were to follow the Levinasian approach to ethics, one would not be interested in trying to formulate some new theory of why or how the machine could or should have a face. So one would not be interested in for instance knowing that everything is “really” information processing, that just as human biology or the human brain, say, the computer is also essentially an information

⁹ See for instance Northup (2014), Elkins (1976), Davis (1966).

processing system. This knowledge, as such, can do nothing for ethics. Instead, what one would have to do is to look, as Wittgenstein would put it, “don’t think but look!” (Wittgenstein 1967, §66), at how one’s relationship to a machine, say a computer, is really fleshed out: the face of the computer, if “it” has one, must bear its meaning by itself — as the face of another — and not gain it through theoretical speculations and abstractions.

It would be unfair to Gunkel to leave unsaid that there are also traits in his writings that strongly suggest this kind of approach — although the articulation of it stays weak and seems to fall back on a centrist logic. As Levinas quite strictly focuses on discussing ethics from the perspective of human otherness, Gunkel and other thinkers eager to include artefacts like machines into the sphere of ethics, have turned to elaborations of Heidegger’s concept of *Gelassenheit*, which, in this context, is usually understood as ‘letting be’ and might also be translated as ‘releasement’ (Introna 2014). Perhaps in more detail than Gunkel — who does take up this discussion — Lucas Introna has dedicated much effort to developing this Heideggerian notion as a way through which to radically break the perceived anthropocentric bias of Levinas’ ethics.

Gelassenheit or ‘letting be’ sets the stage for, what Introna takes to be the “impossible possibility” of opening up to the otherness of all others, humans and non-humans, living beings and the non-living alike. ‘Letting be’ is to ‘dwell’ with beings (which is Heidegger’s account of the Heraclitean concept of *ethos* (Ibid., 50)), to “let the other be as other — that is, as ‘ends’ in itself and for itself”. And as Introna continues “This dwelling is a form of cultivating and care, but what is cultivated, and cared for, is ‘letting be’ — or perhaps more accurately, letting become, exactly as other. Letting be as the dwelling in the midst of the radically other without succumbing to the desire to turn it into something knowable, that is, into something in our image (Ibid.,51). As it stands, this can definitely be seen as taking on the challenge of extending Levinasian ethics to non-human and non-living others, insofar as it does not strive to force ethical content onto e.g. machines but rather suggests that the “dwelling” with machines and letting them be will reveal to us their ethical meaning, as they bear it with themselves. The idea is that by dwelling with things, by letting them ‘touch’ us with their own authentic weight/meaning, the ethical relationship will show itself without any need to impose it from our side — and any attempt to do so would reduce things to means or values.

Although one might appreciate Introna’s and others attempts to develop, through the notion of dwelling with *Gelassenheit*, a relationship with beings that breaks our consumerist and purely instrumental relationship to things — and we can surely be said to be in need of a change in attitude — there are also problems with the notion and the way in which it seems to bring out of focus important considerations. I have at least two central concerns I want to raise here. To begin with, one thing that strikes me as problematic is that in eagerly searching for the ethical relationship to not just the “human other (as Levinas does) but also of more mundane objects such as atoms, hammers, fish, cups, trees and pens” (Ibid., 50), one risks avoiding asking and clarifying the difference between our relationship to living beings and natural things/phenomena on the one hand and artificial things on the other. And the urgency of this distinction, for our moral self-understanding, is that unlike living beings and natural things/phenomena artefacts are in this world and amongst us because we have created them. That is to say, artefacts stand, on the one hand, in relation to human power and control in a specifically different way than do living beings and natural things, and on the other hand artefacts, because they are created by us and do not have an independent nature like living beings do, their meaning is always intertwined within a web of human relations.

In a sense one might of course think of the reproduction of living beings, especially of humans, as being tied to the concept of power, to the concept of a decision to breed. Nevertheless, in reproduction nothing is of course devised and constructed by humans. Genetic engineering should obviously be understood as a technological manipulation of life and in this sense allows us to say that even reproduction can *possibly* be manufactured. But genetic manipulation must of course always presuppose life as such, take for granted that the living organism already is living and has certain characteristics, which then can be manipulated,

controlled, even radically modified etc. An artefact like a machine is not like reproduction, for there is no 'life' already there to be controlled, manipulated etc. *What* the machine becomes, *that* the machine becomes, is, in a sense, absolutely up to humans. When one decides to build a machine, one must decide on *everything*, there is no *given* here — except a tradition of technological know-how; there is no necessary characteristic of what the machine must become like: will it be a computer, steam engine, a robot, a jet engine, will it be something completely new, something useless perhaps? And further critical questions introduce themselves: why build *this* machine, for what purpose, for what reason, on what demand, in what spirit?

To be sure, all of these questions, which are essentially, or as Wittgenstein would say “grammatically” (cf. Wittgenstein 1967, §371-373) connected to artefacts are of course ethical questions in the sense that they are tied to our very relationship to and responsibility for our technology coming to be at all. Our relationship to and responsibility for natural things and phenomena is different. A stone is not the result of my actions and there is no question of why the stone is what it is and why it is that involves my responsibility for it since it is not tied to my aspirations or power. Obviously, we interact with natural things and phenomena all the time and thus play a role in their placing and formations. And to the extent we do this, there is always the question — although perhaps not asked — of why we do it, what drives us to interact with these things and phenomena¹⁰.

So my critical question to Introna, and Gunkel, is: does the dwelling with things and letting them be reveal to us the quality of our desires and aspirations for bringing these artificial things into the world or does it conceal them? Does a thing's own being, as revealed to us in 'dwelling' and 'letting be', reveal to us what 'spirit' it is an expression of? Would a dwelling with a nuclear bomb, letting it be, tell us anything about the dynamics of the self-centred and power driven aspirations that have gone into it, and would it reveal to us how that very violence is part of our everyday and normative social structures and identities? Does a nuclear bomb have an own being independent of these dynamics? How could the 'being' of a technological artefact be detached from the moral dynamics — which is *about* our relationship to each other — that has driven us to construct it?

In one of his texts Introna depicts what the ethical dwelling with a thing could be like (Introna 2014). The example is an old pencil. The description begins with noting the tactical or “fleshy” touch between the pencil and the hand. On this plane the description stays quite abstract, noting that, although in “immediate *contact*”, the hand and the pencil are “in a very real sense also immediately alien to each other — already withdrawing, even in the immediacy and intimacy of the contact” (Ibid., 56). The “fleshy” contact is to be understood as the “primordial affectedness” which is the “condition of possibility” for any human or other perspective, something that “opens the possibility for the radical other to provoke — for the subject to become disturbed, to become its hostage; and as such to become obligated and responsible” (Ibid., 56).

Some critical observations are in place. First, as the description stands, it does not become at all clear whether this “disturbance” and becoming “hostage” clarifies the ethical or whether it rather obscures our understanding and is an expression of a fixation with 'things themselves'. The reason for my suspicion here is exactly that this dwelling does not seem to be sensitive to the distinction between living things and artefacts. “The hand is allowing”, Introna writes, “the pencil to provoke it, in touching it” (Ibid., 56). But could we turn the sentence the other way around, and say that “the pencil allows the hand to provoke it, in touching it”. This does not make any clear sense. What does make sense is that the pencil provokes, *in us*, all kinds of different memories, emotions and thoughts — as Introna does note — exactly because the pencil is of course a 'pencil' insofar as there is a language, insofar as there are practices such as writing, marking, drawing etc. and there are other people or living beings that do such thing, other people we have relationships to. That is to say, all of these practices are already placed within an ethical

¹⁰ The same applies to all kinds of social structuring where humans are ordered and organised under social identities and categories. Elsewhere (Toivakainen 2014 & 2015) I have tried to articulate how social and collective identities form a dynamical relationship to the 'face'. As Levinas writes, the face “is what cannot become a content, which your thought would embrace; it is uncontainable, it leads you beyond (Levinas 1985, 86-87)

space, within a space of meaning, as all these practices are related to networks of relationships between living beings. The pencil, as any artefact, does not have any *independent* character or characteristics, as do living beings, but is an expression — not simply an ‘instrument’ — of human life. One could say that the pencil is on the one hand exactly a ‘thing’, insofar as ‘it’ does not respond in any way, for it is only we who respond to ‘it’. On the other hand one might just as well say that a thing is never just a ‘thing’ because its very coming into being — that it is made — is never simply a relationship between the maker and the ‘thing’ itself, but rather between the maker and the ‘thing’ as an artefact in an already existing space of meanings, in a space of ethics.

If one was to make a pencil one would obviously not make it simply for the ‘thing’ itself — one would not make the pencil *for* the pencil, nor *for* any other artefact — but rather in making it one makes an ‘artefact’, a ‘pencil’, something with a meaning, a meaning that ‘it’, i.e. the pencil, does not give by itself, but rather attains *by being* a ‘pencil’ — a concept¹¹. One might, of course, make a pencil, or any artefact, just for the fun of making it or as a non-functional piece of art. Nonetheless, a thing made for no specific purpose — for to *be* it must be made — has *that* particular meaning precisely insofar as it is situated in an already existing space of meanings, just as a non-functional piece of art is a piece of art because there is already ‘art’.

The example of the pencil is, for reasons stated above, radically different than any “dwelling” with a living being, even with a virus, which is Introna’s example (Ibid., 56). Although it is not at all clear what it would mean to dwell with a virus, the virus nevertheless has its own life and can respond to us, not simply provoke responses *in us* as the artefact does. And this is connected with the concern that the “dwelling” with things risks avoiding what to me seems to be essential aspects of our moral relationship ‘with’ artefacts, namely that it does not critically seem to reflect on and clarify to us the moral dynamics embedded in the aspirations of which artefacts are expressions of. Again the pencil might work as an example. As noted, pencils do not come ‘naturally’ into this world but rather through different human practices, such as marking, writing, drawing etc., in the sense that they are extensions — not simply instruments — of these practices. The point of Levinasian ethics is to say that human life — which can be extended to all living beings — is ethical in the sense that ethics, the ethical relationship to the other, stand at the very constitution of subjectivity, thought and language. When one does a marking, when one writes or draws something one does it within a realm of ethics. One for instance writes love-letters because one yearns for one’s lover. Or, one devotes one’s life to writing novels because one is affected by a trauma and fear close and intimate human relationships and finds writing to ease one’s loneliness. Or, one writes a physics book in order to *share* one’s knowledge in physics, to contribute to the human enterprise of physical sciences. ‘Writing’ is never simply a ‘neutral’ thing one does, but is always intertwined with a moral dynamics. The pencil is part of this dynamics as an extension of the human aspiration to communicate and thus never simply a ‘neutral’ thing: the meaning of the pencil is tied to this dynamics. In other words, as a critical note to Introna, it is not that in our ethical relationship to a ‘thing’, e.g. a pencil, we are responding to ‘it’ or have a responsibility towards ‘it’. For how could the pencil care whether or not it is burned or preserved? Rather, as I would claim, the ethical responsibility in the case of artefacts is a responsibility towards ourselves and the kinds of persons we are. By this I mean that how I relate to a pencil says something about my sensitivity and concern for a communion of beings. But the pencil does not care what, who or how I am, how I treat ‘it’. Living beings on the other hand do care — they respond since I am a face to them — and it is towards them that I am responsible.

To repeat, my suggestion is thus that understanding our ethical relationship to artificial things cannot be detached from a critical examination of what moral dynamics drives our technological aspirations. Such critical examinations are obviously also important for our understanding of contemporary technological development, for instance within artificial intelligence research and robotics. One should, to my mind, for instance stop and reflect on why

¹¹ It is of course important to bear in mind that the descriptions I am giving are not descriptions of a ‘pencil’ as reduced to a purely instrumental ‘thing’. That ‘pencils’ gain their meaning through being artefacts does not *diminish* or *rob* them of anything or reduce them to *mere human purposes*, but is rather part of their ‘grammar’.

our technological development is so strongly pushing for constructing robots and other AI systems that are meant to resemble and simulate human and other animal faces. Are there too few faces so that we are in need of new ones? Have the ‘natural’ faces become too expensive for our economic demands? Or, are we aspiring to construct machines with faces because we are disappointed with the ‘natural’ ones or perhaps because we find artificial ones to be more controllable and thus can avoid the vulnerability of being exposed to living faces and the responsibility of caring for them?

The compensatory logic of the machine-face

As I suggested in the introduction, it is not a coincidence that the debate on the moral status of machines has become so vibrant now that technological developments have started to reach a point where AI systems and robots start interacting with humans in more and more social contexts and on intimate and personal levels. While the question of how we are to understand our ethical relationships to these new artificial intelligence systems is pressing and quite present in both public and philosophical discourses, lesser attention is given to critical examination of *why* we are developing the kinds of technologies that we are. This can be exemplified in one of the concluding remarks made by Gunkel in his book, where he writes:

Should machines like AIs, robots, and other autonomous systems be granted admission to the community of moral subjects, becoming what would be recognized as legitimate moral agents, patients, or both? This question cannot be answered definitively and finally with a simple “yes” or “no.” The question will need to be asked and responded to repeatedly in specific circumstances. (Gunkel 2012, 215)

As can be noted, Gunkel focuses on how we can come to relate to our technological devices, *without* asking why it is that we have these devices and, importantly, why we are aspiring to build ever more sophisticated and life-like machines. When we think of a machine with a face I gather that most of us will be thinking of a robot or some AI system to which we might become attached, towards which we would feel compelled to respond affectively — although Introna’s point is of course that we are always already in an ethical relationship to ‘everything’. To my mind we have already built such machines, a bit in the same sense as we have dolls and other artefacts that come to invoke affection by playing on the responsiveness we have towards human and animal characteristics, which in turn stem from the face. Many of the robots developed, at least for civilian commercial markets, are more or less like dolls or stuffed animals that have an extended ability to evoke responses from humans by including body movement and gestures, sound and sound “recognition”, “memory”, “learning” capacities etc. What then makes such robots different from dolls and stuffed animals is that their “affective computing” diminishes our need to project “life” onto the artefacts since this projection is already, as it were, done on behalf of us (cf. Turkle 2011, 141)¹². But, most robot and AI enthusiasts might say, the dawn of “autonomous” robots and AI systems will change this setting, for then it will not simply be an issue of simulation and projection. Rather, these systems will be more or less self-sufficient and develop their “own” nature, i.e. there will be someone there to respond to. I think that this enthusiasm and fantasy is not so much false or over-optimistic but rather fails to be sensitive to what such aspirations builds on *in us*: what is it that we want to see and understand and what is it we do not want to see and understand — what questions do we avoid asking. Hence I do not feel it to be important to speculate on the limits of technology, on the limits of human power (cf. Ellul 1990, Mumford 2010). As said, my concern here is with why we aspire to devise and construct machines that have a face, when the world is filled with faces.

Probably one of the most discussed and studied contexts in which sociable robots have gained a foothold is eldercare. And it is of course not surprising that eldercare is so exposed to sociable robotics and that it is an issue of public awareness, given the status and state of elderly people in today’s modern society. The issue is obviously complex and I cannot go into it in detail here. Suffice to say though, given that the traditional extended family structure has dissolved in

¹² And here the diminishment should be understood in a double sense: By lessening the need for projection of “life” onto the object, it also lessens the use of our active imagination (quite in a similar manner as a movie displays the landscape that the novel described).

contemporary societies, elderly people find themselves, in growing numbers¹³, lonely — i.e. in need of social companionship — and in need of extra help with everyday tasks and additional care/help in cases of grave functional disability. So the introduction of social robotics seems like a “natural” solution to this societal and interpersonal problem and dilemma¹⁴. Now given this reasoning the project easily and quite automatically becomes one in which we start asking for what is needed for the programming and construction of a care-robot so that it would be able to ‘solve’ the problem: what features does it have to have in order for people to feel that it has a “presence” (Sorel & Draper 2014), what physical characteristics must it have in order for it to be able to, appropriately and displaying features of care, lift a person, help an elderly person move around etc. Obviously these are important questions for the development of these robots, and undeniably such robots might bring much comfort and help to people in need. Nevertheless, the problematic aspect of it all, something very seldom discussed seriously, is that it becomes hard to find room for an understanding of how it is that we have come to a situation in which we need to start developing care-bots in the first place — and what kinds of moral dynamics is involved here — when all resources are focused on ‘progress’ and ‘development’. The point here being, that it is hard to see whether such ‘progress’ stems from the actual ethical meaning others have or whether ethics is forced to accommodate itself to the demands of ‘progress’ — that the demands of ‘progress’ makes ethics *progressive* in the same sense as so called ‘green’ or ‘ethical/fair’ economy makes ethics and ecology economic commodities.

Take for instance the case of the loveable seal-doll Paro¹⁵. One of this seal-doll robot’s main purposes is to ease the loneliness of elderly people and other therapeutic purposes, a “job” that a real live animal might also do. The downside with real animals is of course that, in contrast to Paro, they need real care-taking: one has to feed the animal, wash it, brush it, clean after it etc. Easier and more in accordance with cost benefit demands and lack of resources is a robot seal that can be “put to sleep” at any point, that does not need cleaning up after it etc.¹⁶.

An actual animal used in animal *therapy* is there to compensate for or to fill a relational void — loneliness, depression, disassociation from social and interpersonal life — in the elderly person’s life, a void that is connected essentially to human relations (cf. Turkle 2011). This is of course not to reduce all relationships to animals to compensations for the lack of human relations, but in the context of animal therapy for the elderly, it is quite clear that the need for animals as therapy is connected to a lack of human relationships or, alternatively, to a difficulty of being with humans. Animals — and Paro — can, for the most part, be compensated by humans — for instance by volunteer-workers visiting and spending time with elderly people and/or by increasing the staff of elderly people’s homes — *if* resources allow. However, volunteer-worker and staff-member relationships have an aspect of compensation too insofar as they are formal and institutional ways of securing human contact, when it is lacking. This is obviously not to undermine such work or the genuine encounters it can create, but simply to point out that there is a chain of compensation running from institutionalised human relationships to ‘affective simulations’.

The point of critical examination that increasing technological development of the kind we are now witnessing risks burying under its ability to compensate for absent humans and living beings, is the question how this compensatory logic is bound to for instance our fears and difficulties of being with each other, also during the most unpleasant and difficult times. Are these technologies ways of taking responsibility or avoiding it? Do they build on a conception of freedom as something tied to responsibility, or a conception of freedom solely built on rights and independence from others?

¹³ For more on the statistics of loneliness amongst elderly people see Dykstra (2009).

¹⁴ Elderly people’s loneliness and seclusion is not only a societal problem because most of these people have their own relatives (children, grandchildren) who in turn face the moral dilemma of how to live with the fact that they do not themselves take care of their own parents or grandparents but nevertheless want them to have a decent life.

¹⁵ For more, visit the Paro website at <http://www.parorobots.com/>

¹⁶ As the general manager of an elderly home in Japan (that uses Paro dolls) Taku Kato-ono notes in an interview, animal therapy requires that the animals be taken care of, which the already scarce resources could not possibly allow for. See <http://www.youtube.com/watch?v=PNw4oicWmWU>. Accessed January 4th 2015.

One might of course object here and say that my way of characterising the situation is misconceived or misunderstood: there is something genuinely unique and irreplaceable about the relationship between a human and an artificial other. Attending the “Robo-Philosophy” conference at Aarhus University in August 2014 I tried to find someone that could give me an example that would offer some sign of such indispensability and irreplaceability, a sign that would indicate that sociable robots do not fall back on a compensatory logic. The best case that I came across during that conference was the example of a Korean robot that is used for instructing and teaching children. As it was explained to me¹⁷, one of the main things the robot does with the children is to instruct and guide them through different cognitive games and tasks. As far as I could understand the robot had been and is a great success, since children generally seem to find the robot pleasing and fun to work with and be taught/instructed by. And so I placed my inquiry: was there something irreplaceable, something unique in the relationship between the robot and the children, for as far as I could see, any teacher might have taken the place of the robot just as well. But, I was told, the surprising finding or observation, which, according to my informant, pointed towards the unique and irreplaceable relationship I was inquiring about, was that as studies had shown, when instructed by the robot children tended to learn much better and quicker than with a real human teacher. The explanation for this, I was told, was that in the case of real teachers children tend to get more nervous, have greater stress about expectations which reduces learning capacity etc., while with the robot, they felt more relaxed and playful, perhaps a bit as if the robot was a peer.

An interesting finding indeed, although if one thinks about it, not such a surprising one. For I do not think that it comes as a surprise to anyone that the educational system builds quite strongly on strict hierarchies and high expectations of success and competitiveness — which these same hierarchies guard. In such a constellation it is quite understandable, to my mind, that children come to perceive and even relate to their teachers, and other adults, as judges and score-givers. Could children feel anything but some level of stress in such dynamics of relationships? Anything that breaks this constellation and in some sense flattens the hierarchy will surely generate fresh resources for children.

What is problematic with this example, then, is that here again the robot seems to be ‘solving’ a problem — or contributing with something new and important — when in fact, I would claim, it “obscures a problem by ‘solving’ it without addressing it” (Turkle 2011, 283). Why is our reaction to this Korean robot not that we have important issues to tackle with respect to the education system and more generally, with respect to the pressures of ‘success’ and the power-relations involved in our social structures? Why do we instead wish to understand it as an example of the wonders of technology, of how our new gadgets and machines make things ‘better’, of how such robots are an important token of ‘progress’? Why do we wish to accept the difficulties and corruptness of our social structures and interpersonal relationships and simply construct new “faces” that help us sidestep these difficulties? For *if* robots start taking the place of teachers, we simply “solve” our problems instead of dissolving them. That is to say, instead of seeing through them (or seeing them through) we see past them: we will not have changed (morally speaking), we will just have changed the structure of our social and interpersonal relationships.

Concluding remarks

My point has not to be *essentially* against technology. Rather, the concern I have tried to articulate is that our society and culture is becoming increasingly dominated by a *technological ideology*. I have also tried to indicate the kinds of moral dynamics involved in it and the critical aspects that are being distorted in the pursuit of including machines and artefacts into the ethical universe.

As with any ideology, one of its problems is that it reduces diversity. That is to say, the more our social structures become dominated by technology, or, the more our social

¹⁷ I will simply give the account as it came to me in the discussion. The main importance of this example is not that it is empirically accurate, but rather that it displays a structure of thought and attitudes.

reality is pressured to adopt increasing technological advancements, the more other possibilities and forms of culture and practices start disappearing. Once robots can start taking care of elderly people or teach our children, there is a high risk that we stop critically reflecting on why we need machines to do such things — although, it is fair to say, the reason for the kind of development we are witnessing is of course due to the fact that such critical thinking is already, and has been for a long time, marginalised: technological ideology limits our thinking about other ways of being. Perhaps life would become much more meaningful if our ‘development’ would in fact mean massive decrease of technological dependency — not total independence and in every respect.

It is increasingly important to keep in mind such a simple, yet overshadowed thing as that we are of course not necessarily dependent on modern high technology to care for our older and younger generations. We should also keep in mind that life does not automatically become more meaningful by making it ‘easier’. It is not at all clear why we should for instance have robots that do our cleaning, i.e. what exactly are the forces driving *such* advancements? Obviously robots that do our cleaning make life ‘easier’ in that we do not have to burden ourselves with cleaning up after us, and, it might be argued, give us more quality time with our loved ones. But is it somehow obvious that this is how we should live, that *this* is the way we can come to spend more quality time with our nearest¹⁸. And, isn’t there a critical element of alienation involved here? For in externalising the task of cleaning to a robot, we alienates ourselves from being in touch with, from understanding, exactly what it means to live the kinds of lives we do; there is always someone or something else to take care of the mess our habits cause. It should be noted that a cleaning robot involves an altogether different form of alienation than does, say, the vacuum cleaner. For while the vacuum cleaner aids us with our cleaning, the cleaning robot separates us from the actual act of cleaning. In *this* sense a cleaning robot provides us with mechanism of alienation that bear close resemblance (in certain respects) to slavery or class (cf. Ellul 1976). The same goes for technologies that enable the externalisation of responsibility towards, for instance, our older generation. Obviously a robot that can ‘care’ for our parents and grandparents (and, someday, for us) releases us from the ‘burden’ of doing so — our modern life-styles obviously does not easily include caring for the elderly personally. But can we really come to understand life — and ourselves and others — if we do not live closely and intimately also with those that cannot anymore care for themselves and who are close to dying? Can we understand life without living in an intimate relationship with death, with the death of an other? Robots, to the extent they are meant to take over the task of care work, are extension of the form of alienation that is already included in institutionalised care work. This is not to downplay the welfare state’s services or institutions, but simply to point out that to the extent they are *used* as means to externalise responsibility and builds on fantasies of independence from burdensome and stressing tasks, they are nonetheless forms of alienation.

¹⁸ I would like to point out that anyone who claims that cleaning robots ‘give us more time with our nearest’ should consider at least two essential points. (i) Is the source to our lack of time with our loved ones due to insufficient technologies or to the life-styles we have become accustomed to. (ii) Cleaning is of course not essentially contrary to spending time with loved ones. One can clean together and for instance cook dinner afterwards, spending intensive moments with each other. The problem is not that we have ‘too little time’, but with what we do with the time that we have; how we live our lives and the social structures we abide with.

References

- Backström, J. (2007). *The Fear of Openness*. Åbo: Åbo University Press
- Benso, S. (2000). *The Face of Things: A Different Side of Ethics*. Albany, NY: SUNY Press.
- Berreby, D. (2005). *Us and Them: Understanding your tribal mind*. New York: Little, Brown and Company
- Calarco, M. (2008). *Zoographies: The Question of the Animal from Heidegger to Derrida*. New York: Columbia University Press.
- Darley, J.M. and Batson, C.D., (1973) *From Jerusalem to Jericho: A study of Situational and Dispositional Variables in Helping Behavior*. Journal of Personality and Social Psychology, 1973, vol. 27, No. 1, 100-108.
- Davis, D. B. (1966) *The Problem of Slavery in Western Culture*. Ithaca, NY: Cornell University Press.
- Derrida, J. (2008). *The Animal That Therefore I Am*. Ed. Marie-Louise Mallet. Trans. David Wills. New York: Fordham University Press.
- Dykstra, P.A. (2009). Older adult loneliness: myths and realities. *European Journal of Aging*. Doi: 10.1007/s10433-009-0110-3
- Elkins, S. M. (1976) *Slavery: A Problem in American Institutional and Intellectual Life*, 3rd revised edn. Chicago: University of Chicago Press
- Ellul, Jacques (1990) *The Technological Bluff*, Transl. Bromiley, W. G. Michigan: William B. Eerdmans Publishing Company
- (1976) *The Ethics of Freedom*. Michigan: William B. Eerdmans Publishing Company
- Floridi, L. (2002). On the intrinsic value of information objects and the info-sphere. *Ethics and Information Technology* 4:287–304.
- (2013). *The Ethics of Information*. New York: Oxford University Press
- Gunkel, J.D., (2012). *The Machine Question*. Massachusetts: The MIT Press.
- Hitlin, S., Vaisey, S. (2010) *Handbook of The Sociology of Morality*. New York: Springer
- Hume, D. (1946). *An Enquiry concerning the Principles of Morals*. La Salle: The Open Court Publishing Company
- Introna, L. (2014). *Ethics and the flesh: being touched by the otherness of things*. In Olsen. & Pétursdóttir (eds.), *Ruin memories: material, aesthetics and the archeology of the recent past*. Oxford: Routledge
- Joyce, R. (2001). *The Myth of Morality*. Cambridge, UK ; New York : Cambridge University Press
- Kant, I. (2002). *Groundwork for the Metaphysics of Morals*. Transl. by Wood, A. W. New Haven: Yale University Press
- Levinas, E. (1969). *Totality and Infinity: An Essay on Exteriority*. Trans. Alphonso Lingis. Pittsburgh, PA: Duquesne University.
- (1985). *Ethics and Infinity: conversations with Philippe Nemo / Emmanuell Levinas*; Transl. by Cohen, R. A. Pittsburgh: Duquesne University Press.
- Mumford, L. (2010) *Technics & Civilization*, fourth edition with a new foreword by Langdon Winner. Chicago: University of Chicago Press
- Northup, S. (2014). *12 Years a Slave*. London: William Collins
- Nykanen, H. (2002). *The "I" the "You" and the Soul: An Ethics of Conscience*. Åbo: Åbo Akademi University Press
- (2014). *Conscience and Collective Pressure*. Philosophy, Psychiatry & Psychology. 21:1, 51-56.
- Sorle, T., & Draper, H. (2014). *Robot cares, ethics and older people*. Ethics and Information Technology 16:183-195.
- Toivakainen, N. (2014). *Social Robots as Mirrors of (Failed) Communion*. In Seibt, J., Hakli, R., Nørskov, M. (Eds.). *Sociable Robots and the Future of Social Relations*. Volume 273 of Frontiers in Artificial Intelligence and Applications. IOS Press
- (2015). *The Moral Roots of Conceptual Confusion in Artificial Intelligence Research*. (Forthcoming) American Philosophical Association Newsletter *Philosophy and Computers*, vol. 14, No. 2
- Turkle, S. (2011). *Alone Together*. New York: Basic Books.
- Wallgren, T. (2006) *Transformative Philosophy*. New York: Lexington Books
- Westerlund, F. (2014). *Heidegger and the problem of Phenomenality*, Helsinki: Philosophical Studies from the University of Helsinki, Unigrafia.
- Wittgenstein, L. (1967). *Philosophical Investigations*. Transl. by Anscombe, G.E.M. Oxford: Blackwell
- (1980). *Culture and Value*. Transl. by Winch, P. Chicago: The University of Chicago Press